Working with Whole Genome Data

Beren Jafar

Department of Bioinformatics, University of Wollongong.

Abstract

The ability to create high-quality sequence data in a public health laboratory enables the identification of disease-causing strains, the strong formal decision about something of relatedness among sudden start of something bad like disease strains, and the analysis of genetic information related disease causing agent and antimicrobial-resistance genes. However, the analysis of whole sequence data depends on bioinformatics analysis tools and processes. Many public health laboratories do not have the bioinformatics abilities to carefully study the data created from putting in correct order and therefore are unable to take full advantage of the power of whole sequence data putting in correct order. The goal of this way of seeing sensible view of what is and is not important is to provide a guide for laboratories to understand the bioinformatics analyses that are needed to understand whole sequence data and how these in silico analyses can be put into use in a public health laboratory setting easily, affordably, and, sometimes, without the need for intensive calculating valuable supplies and basic equipment needed for a society to operate.

Keywords

Bioinformatics, Next-generation sequencing, Genes, Genome, Genomic data

Correspondence to:

Beren Jafar,
Department of Bioinformatics
University of Wollongong.
E-mail: jafarbe34@gmail.com

Citation: Jafar B (2021).Working with Whole Genome Data. EJBI. 17(7): 1-2
DOI: 10.24105/ejbi.2021.17.7.01-02
Received: July 03, 2021
Accepted: July 17, 2021
Published: July 24, 2021

1. Introduction

Next-generation sequencing (NGS), also known as high-throughput data sequencing, has affected many fields in the study of qualities of living things but has changed a lot the field of genomics by enabling people who work to find information to quickly sequence whole microbial genomic data, profile genetic expression by putting in correct order RNA, examine host-foreign particle interactions, and study the huge microbial diversity in humans and the surrounding conditions (Koboldt DC et al. [1]). Even though there is the existence of the benefits of NGS over traditional Sanger sequence methods, public health laboratories (PHLs) have been slow to put into use this technology. For laboratory secretly recording foodborne sicknesses, pulse-field gel electrophoresis (PFGE) is now the preferred method for typing bacterial isolation and is widely used to find out the proper source tracking. PFGE has been the most important part of the success of CDC‘s PulseNet program since 1997 (Swaminathan B et al. [2]). By 2018 PulseNet program is try to change PFGE with WGS This arc-like path looks like the path taken in the study of human whole genome, in which genetic mapping based on restriction fragment length polymorphism was replaced by almost-complete information received by high-throughput whole genome putting in correct order. Although restriction fragment length polymorphism markers at first enabled the measurement of genome distance and laid the foundation for linkage mapping, its success depended on said phenotypic effects of the hidden feature and regularly broke up and moved away markers. Once linkage to an area was identified, things causing other things to happen could be clearly identified through fine mapping. WGS given not only a complete marker-map with maximum ability to display at the nucleotide level but also enabled the deduction of things causing other things to happen and direct testing of genome relatedness and genomic origination beginning. The promise of this approach also extended to the study of disease causing agents, given that WGS in the end enables testing of clearly stated guesses looking at genotype-phenotype relationships e.g., antimicrobial drug resistance. However, although more PHLs are adopting NGS and WGS, only a small number of these laboratories have the ability to do the bioinformatics analyses needed to take full advantage of the data they are creating. CDC aids PHLs in conducting foodborne disease observation on a national scale but is unable to help with data analysis for local foodborne disease observation. Some of the stopping things preventing PHLs from putting into use the bioinformatics-dependent analysis are the needed things for large-scale computer-based abilities, complex molecular transformative or changing analyses, and dedicated bioinformatics staff to do these analyses. However, it‘s almost necessary to need a computer with proper internet connection. Many of these tools are open-source and can be used to do a range of bioinformatics analyses. Two of these tools are Illumina‘s Base Space Sequence Hub (Illumina, Inc., San Diego, CA, USA) and the web-based forum. Analysis of sequencing data through Bioinformatics is often done in a multistep, with 1 input and 1 output process. It consists of 8 steps like, read quality control, reference strain strong formal decision about something, read mapping to the reference strain, single-nucleotide polymorphism and small insertion or deletion detection, de novo genome sequencing, genome interpretation, phylogenetic tree construction, and phylogenetic analysis. Although such processes are standard, multiple software solutions are available for these steps. These multistep, multi-software analyses are often set up to run automatically from 1 step to the next without input from the user (Bolger AM et al. [3]).

Conclusions

Many years have passed since the release of human reference genetic sequencing. With the moving ahead of the genome assembly technology, hundreds and thousands of whole-genome can be gotten in single institute within a short period. Also, WGS data analysis applications, including hardware and software-based solutions, would speed up to allow large-scale data analysis on multi-cloud by combining their dataset to available human genetic data with population scale via their data sharing policy. Therefore, in human genome, from the outputs of the workflow engines to the large-scale human genomic data, more domain-specific downstream data understanding would be demanded from both the expert-knowledge driven approach by the domain knowledge from the medical, professional biologists and the data-driven approach from computer science, e.g., intelligent retrieval.

References

1. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. Cell. 2013; 155:27–38.
2. Swaminathan B, Barrett TJ, Hunter SB, Tauxe RV. CDC PulseNet Task Force. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg Infect Dis. 2001.
3. Bolger AM, Lohse M, Usadel B Bioinformatics. 2014; 30(15): 2114-2120.